

Using grammar checkers in the ESL classroom: the adequacy of automatic corrective feedback

Paul John¹ and Nina Woll²

Abstract. Our study assessed the performance of two Grammar Checkers (GCs), Grammarly and Virtual Writing Tutor, and the grammar checking function in Microsoft Word on a broad range of grammatical errors. The errors occurred in both authentic English as a Second Language (ESL) compositions and simple sentences we generated ourselves. We verified the performance in terms of (1) coverage (rates of error detection), (2) accuracy of proposed replacement forms, and (3) ‘false alarms’ (forms mistakenly flagged as incorrect). To the extent GCs provide accurate and comprehensive corrective feedback, they could relieve teachers of the time-consuming task of providing written feedback themselves. While inaccurate replacement forms and false alarms are relatively rare, we found GCs to have poor overall coverage (total error detection rates under 50%). Grammarly and Virtual Writing Tutor, however, outperform Microsoft Word. Coverage is also higher both for certain categories of error and for the sentences rather than the authentic compositions. Finally, although GCs do not provide comprehensive feedback, we suggest designing special activities that target select error types.

Keywords: grammar checkers, corrective feedback, focus-on-form, second language learning.

1. Introduction

Our study investigates the adequacy of automatic corrective feedback from GCs to determine their possible use in the ESL classroom. Written corrective feedback permits teachers to incorporate a focus on form into the communicative classroom, thereby promoting accuracy and preventing fossilization (Bitchener, 2008; Ferris,

1. Université du Québec à Trois-Rivières, Trois-Rivières, Canada; paul.john@uqtr.ca

2. Université du Québec à Trois-Rivières, Trois-Rivières, Canada; nina.woll@uqtr.ca

How to cite this article: John, P., & Woll, N. (2018). Using grammar checkers in the ESL classroom: the adequacy of automatic corrective feedback. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouéšny (Eds), *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 118-123). Research-publishing.net. <https://doi.org/10.14705/rpnet.2018.26.823>

Liu, Sinha, & Senna, 2013). Still, providing feedback is time-consuming, so the potential for GCs to relieve teachers' workloads is appealing. In essence, GCs look like an invaluable tool for the ESL context.

Important questions remain, however, concerning the quality of automatic corrective feedback. Previous studies have often adopted a narrow focus, evaluating GCs only on articles/determiners, prepositions, and collocations (De Felice & Pulman, 2008; Han, Chodorow, & Leacock, 2006). Research on the grammar checking function in automated writing evaluation systems has been more comprehensive (Dikli & Bley, 2014, on Criterion), but these systems are prohibitively expensive. In our view, an investigation of GCs available for little or no cost and on a wide range of grammatical issues is overdue. The current study thus addresses the following research questions:

- To what extent is automatic corrective feedback comprehensive and accurate?
- Do GCs perform better on certain grammar points than others?

2. Method

2.1. Data collection

We evaluated two leading online GCs (Grammarly and Virtual Writing Tutor) and the grammar checking function in Microsoft Word on errors from two sources: (1) authentic compositions (50 handwritten essays generated under exam conditions by 28 francophone TESL³ students at a university in Quebec; 10M /18F; age 21-36); and (2) a set of 129 simple sentences containing errors we generated based on our knowledge of typical francophone errors.

Representative errors were selected from the compositions, and these errors and the simple sentences were run through the three GCs to verify coverage (error detection rates) and accuracy of proposed replacement forms. The 50 compositions and 129 sentences were then run through the GCs to establish rates of 'false alarms' (forms mistakenly flagged as incorrect).

3. Teaching English as a second language.

2.2. Results

Table 1 shows how the GCs performed on the two sets of errors (compositions vs. simple sentences) in the different grammatical categories listed on the left. The results are presented as fractions, such that 2/4, for example, indicates that the GC identified two out of four errors (Gram=Grammarly; VWT=Virtual Writing Tutor). Though many of the error categories are self-evident, others may be elusive. By ‘tense shift’, we mean shifts primarily between past and present in contexts where either is acceptable. The category ‘plural nouns’ refers to failure to pluralize a noun or pluralization of a non-count noun. Possessive errors involve inappropriate use of either apostrophe + ‘s’ or the periphrastic possessive with ‘of’. Pronoun errors concern incorrect reference. The category ‘relative clauses’ refers to incorrect comma usage with restrictive and non-restrictive relative clauses.

Table 1. Rates of error detection: compositions vs. simple sentences

	Grammatical categories	Compositions			Sentences		
		Word	Gram	VWT	Word	Gram	VWT
Verbs	Tense-aspect	2/4	1/4	2/4	1/9	4/9	0/9
	Verb form	1/3	3/3	2/3	2/13	8/13	8/13
	Subj-V agreement	0/3	3/3	0/3	0/6	6/6	6/6
	Tense shift	0/6	0/6	0/6	0/2	0/2	0/2
	Total	3/16	7/16	4/16	3/30	18/30	14/30
Nouns	Plural	1/3	3/3	3/3	4/20	11/20	11/20
	Possessive	0/5	3/5	2/5	0/4	0/4	0/4
	Pronoun	0/2	0/2	0/2	0/5	2/5	0/5
	Total	1/10	6/10	5/10	4/29	13/29	11/29
Preps	Wrong prep	0/3	1/3	1/3	0/10	8/10	8/10
	Missing prep	0/2	0/2	0/2	0/4	2/4	2/4
	Unnecessary prep	0/2	0/2	0/2	0/7	3/7	2/7
	Total	0/7	1/7	1/7	0/21	13/21	12/21
Words	Word order	0/3	0/3	0/3	3/18	7/18	3/18
	Word form	0/3	0/3	0/3	6/10	7/10	7/10
	Total	0/6	0/6	0/6	9/28	14/28	10/28
Misc.	Determiner	0/4	0/4	0/4	1/13	4/13	4/13
	Relative clause	0/3	0/3	0/3	2/8	1/8	0/8
	Total	0/7	0/7	0/7	3/21	5/21	4/21
Grand totals		4/46 (8.7%)	14/46 (30.4%)	10/46 (21.7%)	19 (14.7%)	63 (48.8%)	51 (39.5%)

The grand totals in Table 1 indicate poor overall error detection (all below 50%). In addition, Microsoft Word achieves considerably lower coverage than the two

online GCs, with Grammarly generally outperforming Virtual Writing Tutor: hence, Grammarly >> Virtual Writing Tutor >> Microsoft Word. Error detection is greater on simple sentences than on compositions. In addition, there are some grammatical categories in which Grammarly, and to a degree Virtual Writing Tutor, perform better: particularly verb forms, subject-verb agreement and plural nouns. They are also strong in the ‘wrong preposition’ and ‘word form’ categories, but only with simple sentences. Finally, we can report that incorrect replacement forms are rare: we found one inaccurate replacement for Grammarly, three for Virtual Writing Tutor and four for Microsoft Word.

While none of the GCs raised false alarms in the simple sentences, Grammarly shows a clear edge over both Virtual Writing Tutor and Microsoft Word for false alarms on the compositions (see [Table 2](#)). The absence of false alarms on the simple sentences is partly due to lack of opportunity (1,055 words in the sentences vs. 23,108 words in the compositions). Microsoft Word’s relatively low number of false alarms is probably a function of its low rate of error detection.

Table 2. Rates of false alarms

	Microsoft Word	Grammarly	Virtual Writing Tutor
Compositions	13	4	30
Simple sentences	0	0	0

3. Discussion

We evaluated the performance of two online GCs, Grammarly and Virtual Writing Tutor, and the grammar checking function in Microsoft Word on a wide range of grammatical errors. The fact that Grammarly and Virtual Writing Tutor clearly outperform Microsoft Word in error detection suggests that learners should be wary of relying on this omnipresent word processor to check the accuracy of their writing. They might instead consider turning to an online GC for a fuller picture.

Nonetheless, Grammarly and Virtual Writing Tutor also show limited coverage – which parallels the findings in [De Felice and Pulman \(2008\)](#) and [Han et al. \(2006\)](#). An important implication is that ESL teachers cannot truly count on the technology to provide comprehensive written corrective feedback on student compositions. The fact that error detection rates were higher for the simple sentences than for the authentic compositions simply underscores this conclusion.

The low rates of inaccurate replacement forms and false alarms are encouraging for the ESL context. Inaccurate feedback could lead ESL learners seriously astray, particularly since they lack native speaker intuitions to override misleading feedback. It is encouraging that GCs perform strongly in some categories of error (verb forms, subject-verb agreement, plural nouns, wrong prepositions, and word forms). We suggest that teachers use GCs to target specific error types in student compositions and encourage students to scrutinize their own writing for errors that the GC might have overlooked. Furthermore, teachers can develop special activities containing errors that the GCs are capable of identifying. Students can first try to identify the errors themselves and then run the text through the GC to check their answers.

4. Conclusions

While our findings show that GCs have poor overall coverage, Grammarly and Virtual Writing Tutor have higher coverage than Microsoft Word. GCs are also better at detecting errors in some categories than others and in specially composed simple sentences than in authentic compositions. Finally, both inaccurate replacement forms and false alarms are infrequent. Thus, though GCs cannot provide comprehensive corrective feedback on student compositions, they can be employed to target select error types in student writing and in specially developed activities alike. In this manner, GCs can be used effectively to incorporate a focus on form into the communicative ESL classroom.

5. Acknowledgements

We appreciate the invaluable input of our colleagues, Mariane Gazaille and Walcir Cardoso, and research assistant, Michel Monier.

References

- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17, 102-118. <https://doi.org/10.1016/j.jslw.2007.11.004>
- De Felice, R., & Pulman, S. G. (2008). A classifier-based approach to preposition and determiner error correction in L2 English. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, 169-176. <https://doi.org/10.3115/1599081.1599103>

- Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: how does it compare to instructor feedback? *Assessing Writing*, 22, 1-17. <https://doi.org/10.1016/j.asw.2014.03.006>
- Ferris, D., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing*, 22, 307-329. <https://doi.org/10.1016/j.jslw.2012.09.009>
- Han, N., Chodorow, M., & Leacock, C. (2006). Detecting errors in English articles usage by non-native speakers. *Natural Language Engineering*, 12(2), 115-129. <https://doi.org/10.1017/S1351324906004190>

Published by Research-publishing.net, a not-for-profit association
Contact: info@research-publishing.net

© 2018 by Editors (collective work)
© 2018 by Authors (individual work)

Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018
Edited by Peppi Taalas, Juha Jalkanen, Linda Bradley, and Sylvie Thouësny

Publication date: 2018/12/08

Rights: the whole volume is published under the Attribution-NonCommercial-NoDerivatives International (CC BY-NC-ND) licence; **individual articles may have a different licence.** Under the CC BY-NC-ND licence, the volume is freely available online (<https://doi.org/10.14705/rpnet.2018.26.9782490057221>) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.

Disclaimer: Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it was not under consideration for publication elsewhere. While the information in this book is believed to be true and accurate on the date of its going to press, neither the editorial team nor the publisher can accept any legal responsibility for any errors or omissions. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

Trademark notice: product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Copyrighted material: every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net
Cover theme by © 2018 Antti Myöhänen (antti.myohanen@gmail.com)
Cover layout by © 2018 Raphaël Savina (raphael@savina.net)
Drawings by © 2018 Linda Saukko-Rauta (linda@redanredan.fi)

ISBN13: 978-2-490057-22-1 (Ebook, PDF, colour)

ISBN13: 978-2-490057-23-8 (Ebook, EPUB, colour)

ISBN13: 978-2-490057-21-4 (Paperback - Print on demand, black and white)

Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, UK: British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: Décembre 2018.